

引用文献 |

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-168868

(P2002-168868A)

(43)公開日 平成14年6月14日 (2002.6.14)

(51)Int.Cl'	識別記号	F I	マーク*(参考)
G 01 N 33/58		G 01 N 33/58	A 2 G 0 4 3
C 12 M 1/00		C 12 M 1/00	A 2 G 0 4 5
C 12 N 15/09		C 12 Q 1/68	Z 4 B 0 2 4
C 12 Q 1/68		G 01 N 21/64	F 4 B 0 2 9
// G 01 N 21/64		C 12 N 15/00	A 4 B 0 6 3

審査請求 未請求 請求項の数3 OL (全6頁)

(21)出願番号 特願2000-362648(P2000-362648)

(22)出願日 平成12年11月29日 (2000.11.29)

(71)出願人 000001993

株式会社島津製作所

京都府京都市中京区西ノ京桑原町1番地

(72)発明者 狹間一

京都府京都市中京区西ノ京桑原町1番地

株式会社島津製作所内

(74)代理人 100065464

弁理士 野口 雅雄

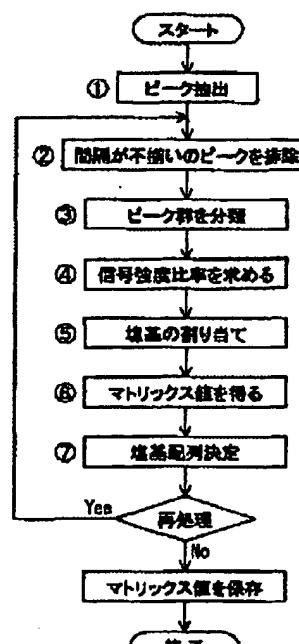
最終頁に続く

(54)【発明の名称】 塩基の塩基配列決定方法

(57)【要約】

【課題】 専用試薬キットを用いることなく、実際のサンプル泳動からマトリックス値を得る。

【解決手段】 泳動波形から信号の出始めの一一定範囲においてピークを抽出し、間隔の揃っていないピークを排除する。ピーク群を信号の強度から分類し、分類された4グループの信号強度比率を得る。分類された4グループに対して、対応する塩基を割り当て、各塩基グループのピーク波形の信号強度比によりマトリックス値を得る。そのマトリックス値を用いて塩基配列決定を行なう。



(2)

特開2002-168868

2

1

【特許請求の範囲】

【請求項1】 複数の異なる蛍光波長を持つ蛍光色素を用いる蛍光色素ターミネーターラベル法で、蛍光色素ごとの検出部から得られた波形信号にマトリックス変換を行なって塩基ごとの信号波形を求め、それに基づいて塩基配列決定を行なう方法において、前記マトリックス変換を行なうためのマトリックス値を実際のサンプル泳動から以下の工程により得ることを特徴とする核酸の塩基配列決定方法。

- ① 適当な範囲からピークを抽出する工程、
- ② ピーク間隔の揃っていないピークを排除する工程、
- ③ 信号の強度に基づいてピーク群を塩基の種類に対応する4グループに分類する工程、
- ④ 分類された4グループの信号強度比率を求める工程、
- ⑤ 分類された4グループに対して、対応する塩基を割り当てる工程、
- ⑥ 各塩基グループのピーク波形の信号強度比によりマトリックス値を得る工程。

【請求項2】 得られたマトリックス値を用いて塩基配列を決定した後、その配列決定した塩基のピーク信号を用いて再度マトリックス値を求める請求項1に記載の塩基配列決定方法。

【請求項3】 条件を限定しておくことにより上記工程のうちの少なくとも1つの工程においてその処理を簡略化する請求項1又は2に記載の塩基配列決定方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はDNAなどの核酸の塩基配列を決定するシーケンシングに関するものである。

【0002】

【従来の技術】 DNAの塩基配列は、塩基ごとに異なる蛍光色素で標識されたDNA断片試料を電気泳動させ、4種類の波長を各々選択的に検出する4種類の検出部で得られた信号ピークの強度（高さ）を基に決定される。蛍光色素ターミネータにおけるdRhodamin色素の規格化された発光スペクトルを図2（“ABIPRISM (Applied Biosystems社の登録商標) BigDye (Applied Biosystems社の登録商標) Terminator Cycle Sequencing Ready Reaction Kit”から引用）に示す。4種類の検出部は4種類の蛍光色素（dR110, dR6G, dTAMRA, dROX）それぞれを最も感度よく検出するように設定されている。

【0003】しかし、各蛍光色素のスペクトルは決してシャープなものではなく、左右の検出部にも、そのすそ野部分が当然もれて検出される。例えば、dR6Gで標識されている塩基A（アデニン）のピーク波形は、図3に示されるように、強度の差はあるが、4種類の蛍光色素用の検出部すべてで検出される。この時の信号強度比率（Pa : Pt : Pg : Pc）は一定なので、この値を元に逆変換

ことになる。これは他の3種類の蛍光色素についても同じである。

【0004】すなわち、検出部での信号強度を次のように表わす。

塩基Aのピーク波形の信号強度比率=APa (=1) : Apg : Apc : APt

塩基Gのピーク波形の信号強度比率=Gpa : Gpg (=1) : Gpc : GPt

塩基Cのピーク波形の信号強度比率=CPa : Cpg : Cpc (=1) : Cpt

塩基Tのピーク波形の信号強度比率=TPa : Tpg : Tpc : TPt (=1)

塩基Aによる発光強度=Ia

塩基Gによる発光強度=Ig

塩基Cによる発光強度=Ic

塩基Tによる発光強度=It

塩基A用の検出部Daで検出される信号強度=Oa

塩基G用の検出部Dgで検出される信号強度=Og

塩基C用の検出部Dcで検出される信号強度=Oc

塩基T用の検出部Dtで検出される信号強度=Ot

【0005】このとき、蛍光色素による発光強度(Ia, Ig, Ic, It)と受光信号強度(Oa, Og, Oc, Ot)との間には、次の行列で表わされる関係がある。

【数1】

$$\begin{bmatrix} Oa \\ Og \\ Oc \\ Ot \end{bmatrix} = \begin{bmatrix} APa & GPa & CPa & TPa \\ APg & GPg & CPg & TPg \\ APc & GPc & CPc & TPc \\ APt & GPt & CPt & TPt \end{bmatrix} \begin{bmatrix} Ia \\ Ig \\ Ic \\ It \end{bmatrix}$$

↑ 行列M

【0006】したがって、得られた信号波形(Oa, Og, Oc, Ot)から元信号、すなわち塩基(蛍光色素)ごとの信号波形(Ia, Ig, Ic, It)を得るには、上の式の両辺に行列Mの逆行列をかけねばよい。この逆行列がマトリックス値である。ピーク信号が別の塩基と重なっている場合にも、検出される波形は単なるスペクトルの加算と考えられる。

【0007】したがって、4種類の蛍光色素についての

信号強度比率が得られれば、それを行列に表し、その逆行列を元の検出ピーク波形にかけねば、各蛍光色素（4種類の塩基）のピーク波形が得られる。この信号強度比率を正確に求めることができ、蛍光色素のマトリックス値を求めることがある。蛍光色素のマトリックス値を得る方法は、塩基ごとに異なる蛍光色素で標識された塩基を一種類ずつ泳動させ、4種類の波長を各々選択的に検出する4種類の検出部で得られた信号ピークの強度（高さ）を計測して行なうのが、一般的である。

【0008】

(3)

特開2002-168868

3

ス値は、各塩基を標識する蛍光色素や、光学系を含む信号検出系によってある種固有のものなので、泳動・検出を行なうハードウェアを調整、あるいは部品交換を行なった際には、その都度新しい値を設定する必要がある。その反面、一度設定すれば、何らかの理由で不具合が発生する（マトリックス値がずれる）までは再設定する必要はない。

【0009】蛍光色素ターミネーターラベル法のための試薬キットは、最初から4種類の蛍光色素がミックスされているので、マトリックス値キャリブレーション用には、蛍光色素が別々になっている特別な試薬キットが必要となる。

【0010】さらに、この専用試薬キットを用いてキャリブレーションのための泳動をする必要がある。これは、ルーチン的に泳動を行なう場合には、最初だけなので問題にはなりにくいが、泳動系の条件を変える実験をしているとか、試薬キットそのものの評価実験をしていくとか、光学系の調整を繰り返している場合等には、はなはだ面倒で、コストのかかる作業となる。

【0011】専用試薬キットを用い、塩基ごとに異なる蛍光色素で標識して、一種類ずつ泳動しなければならないのは、まず、得られるピーク波形がどの塩基のものか明示的に区別がつかないためである。次に、ピーク波形について最も信号強度がある検出部をその塩基と仮定する（例えば、図3の場合には塩基A）手法を用いても、蛍光色素ごとの移動度の差から、そのピーク波形が一塩基からだけなっており、決して一部分たりとも他の塩基と重なってはいない、という保証はないためである。図4に概略的に示されるように、例えばグアニンGの移動度がアデニンAの移動度より大きいとすれば、ピークAとGが部分的に重なることは起こりうる。そして、部分的に重なっていれば、信号強度比率が変化し、正確なマトリックス値が得られなくなる。

【0012】以上の問題を解決するには、純粧に一塩基だけからなるピーク波形を、塩基毎に抽出できる方法を見つければよい。本発明はこれを実現し、専用試薬キットを用いることなく、実際のサンプル泳動からマトリックス値を得る方法を提供することを目的とするものである。

【0013】

【課題を解決するための手段】以下に請求項1の方法を図1を参照して順に説明する。

①ピークを抽出する。泳動波形は、信号の出始め部分でS/Nのよいきれいなピーク波形が得られることが多いので、まず、信号の出始めの一一定範囲におけるピークの抽出から始める。この際のピークの基準は、最も大きな蛍光色素信号の強度が、使用するベースコーラ（塩基配列決定用プログラム）でピーク検出する最低水準より大きいことである。これは小さな信号ではS/Nが劣化するか

4

【0014】②ピーク間隔の揃っていないピークを排除する。移動度の遅い塩基と速い塩基が連続する塩基配列ではピークが重なることが多く、この場合、ピーク間隔が前後で不均一になる。このような部分を検出し、そのピークを排除すれば、移動度の問題の大部分を解決できる。

【0015】③ピーク群を信号の強度から分類する。例えば、BigDyeターミネータにおいて、A（アデニン）グループ [Pa>Pt>Pg>Pc]、T（チミン）グループ [Pt>Pa>Pc>Pg]、G（グアニン）グループ [Pg>Pa>Pt>Pc]、C（シトシン）グループ [Pc>Pt>Pa>Pg] 等である。本来なら4塩基があるので4種類に分類されるはずであるが、サンガー反応や精製の失敗、あるいはノイズ等の問題から、他にも分類されるピーク群が出現する可能性がある。この場合、そのような異常なピークは出現頻度が少ないという前提で、ピーク数の多い順に上位4グループの分類を選択する。また、隣り合う波長の蛍光色素の信号強度より、離れた波長の蛍光色素の信号強度が強いといった場合にも、異常としてそのピークを排除する。これらの排除処理によって、②で排除できなかつた移動度の差から生ずる重なりピークをさらに除ける。

【0016】④分類された4グループの信号強度比率を得る。グループ毎に計算して求める。計算方法は、平均値、中央値等いろいろと利用できる。

【0017】⑤分類された4グループに対して、対応する塩基を割り当てる。信号強度比は、ピークがA（アデニン）の時にはPaが最も強く、T（チミン）の時にはPtが最も強いというのが本来である。しかし、検出器の感度設定等により、逆転していることもある。例えば、BigDyeターミネータにおいて、A（アデニン）のピークなのに、アデニンの検出器の感度が悪いかチミンの検出器の感度が良いかして、[Pt>=Pa] となって、T（チミン）のように見える場合である。ただし、図5に示されるように、A（アデニン）が [Pt>=Pa>Pg>Pc] となっており、T（チミン）が [Pt>=Pa>Pc>Pg] となっている場合には、3番目に大きな信号で区別がつくので、隣の波長であるPgからA（アデニン）だと判る。どちらも [Pt>=Pa>Pg>Pc] である場合には、2つのグループについて隣の波長であるPgの強度比率Pg/Paを比較し、大きい方のグループをA（アデニン）とする。

【0018】⑥各塩基グループのピーク波形の信号強度比によりマトリックス値を得る。塩基が割り当てられた各グループのピーク波形の信号強度比を求め、その信号強度比の行列を作成する。その行列の逆行列を計算してマトリックス値を得る。

【0019】⑦通常のベースコール（塩基配列決定）を行なう。得られたマトリックス値を用いて波形信号にマトリックス変換を行なって、塩基ごとの信号波形を求め、それに基づいて塩基配列を決定することができる。

(4)

特開2002-168868

6

5

マトリックス値を得る。ベースコーラは一般的に、配列決定した塩基に対して、その信頼度的な重み付けを行なう。このステップでは、ほぼ確実に正解であるという重み付けをされた塩基（ピーク信号）を、全データ範囲について抽出し、その波形信号情報を用いて②から④を再度行なう。ここで処理する対象のピーク群は、①で得られたピーク群よりも、信号波形としておしなべて優れており、さらにデータ範囲も信号の出始めだけではなく広範囲にわたっているためピーク数も多い。したがって、さらに精度が高く正確なマトリックス値が算出される。

【0021】④得られたマトリックス値を保存する。次からはこのマトリックス値を用いてベースコールを行うことになる。この場合、泳動条件や試薬キット毎に異なるインデックスをマトリックス値に付加しておけば、それを呼び出すことでマトリックス値の振り分けが簡略化される。

【0022】もちろん、ターゲットとするサンプル泳動が、サンガー反応や精製の失敗、あるいはポリマーとゲルのトラブル、さらにはノイズの問題等でマトリックス値を作成できない場合がある。例えば④では、分類されるべき上位の4グループが他のグループとその内包するピーク数において、明確な差が得られないなど、また④では、正解であるという重み付けをされた塩基（ピーク）の数が少ないとある。特に④において、頼るべきピークを多く得られなかったということは、もとのマトリックス値が間違っている可能性が大である。もとよりこののようなサンプル泳動は、マトリックス値を得るためにターゲット（原本）に用いてはならない。

【0023】上記の方法においては、諸条件を限定せずに一から行なう方法を述べたが、以下に請求項3に対応して諸条件を限定した簡単な方法についての説明を行なう。限定すべき諸条件は以下の2点である。

(1) 検出部の感度を、ピークがA(アデニン)の時にはPaが最も強く、T(チミン)の時にはPtが最も強く、G(グアニン)の時にはPgが最も強く、C(シトシン)の時にはPcが最も強いというように設定する。この調整は、各塩基の信号強度が結果的に揃ってしまうという副次的効果を生むことが多い。これはベースコーラにとって非常に好ましいことなので、逆説的に言えば、各塩基のピーク高さを揃えるためには、少しばかり強度の逆転が起こつても、本方法では構わない。

(2) 反応試薬キットによって、蛍光色素間の移動度の差や強度差がわかっているので、これを予めアルゴリズムに埋め込む。

【0024】項目(1)は、S/N(信号対ノイズ比)を確保し、精度の高い泳動を行なうために、泳動系の調整時に本来すべき基本的な内容である。また、項目(2)においても、泳動毎に全く性格の異なった反応試薬キットを用いることはなく、既存の反応試薬キットを急頭において、ベースコーラを作成。シーケンシングオフのが通常

であり、それがベースコーラの精度を上げているのである。つまり、項目(1)も項目(2)も限定すべき諸条件というよりは、DNAシーケンシングシステムにおいて確度高い配列決定を行なうための当たり前の処置なのであって、決して大きな負担となるものではない。

【0025】蛍光色素毎に検出部で検出される信号強度の順番と比の傾向は、項目(1)の感度調整と項目

(2) の蛍光色素間の強度差がわかっていてれば、おおまかにせよある程度断定的に予測でき、抽出されたピークを始めから決めうちして4種類の塩基に分類するにはその程度で十分なのである。したがって、これにより③と⑤の処理内容を大幅に減ずることができる。

【0026】また、項目(2)から蛍光色素の移動度がわかっていてれば、②のピーク選別において、ピーク間隔のばらつきが容易に予測でき、取捨選択すべきピークの強度が上がる。例えば、BigDyeターミネータにおいては、G(グアニン)の移動度が最も速いので、G(グアニン)の前後のピークが同じG(グアニン)でない限り、ピーク間隔はG(グアニン)の前側の方が後ろ側に比べて極端に狭くなるといった傾向にある。これは決してピーク間隔の異常ではなく、正常な状態なのである。この場合、前側のピーク間隔が狭すぎて信号強度比率に影響が出るといった状況でない限り、このG(グアニン)のピーク信号は有効なのである。

【0027】

【実施例】請求項3に記載した方法について一実施例を説明する。反応試薬キットはETターミネータ(amerham pharmacia biotech社の登録商標)である。ETターミネータは、移動度についてはT(チミン)だけが少し遅く、他の3塩基の移動度はほぼ同じと見てよい。また、蛍光色素の発光波長は、短波長側からG(グアニン) < T(チミン) < A(アデニン) < C(シトシン)の順である。検出部の感度調整は、各塩基のピーク強度を揃えることを第一にして、強度の僅かな逆転は許容している。

【0028】以下に手順を示す。

[1] ピークの抽出

信号の出始めから50bp(塩基)程度の範囲において、4種類の塩基(信号ピーク)を抽出する。

[G(グアニン)のピーク抽出]

40 A(アデニン)やC(シトシン)より大きく、T(チミン)の90%の強度より大きいピークを、G(グアニン)のピーク候補として抽出する。

[T(チミン)のピーク抽出] A(アデニン)やG(グアニン)の90%の強度より大きいピークを、T(チミン)のピーク候補として抽出する。

[A(アデニン)のピーク抽出] G(グアニン)より小さく、T(チミン)やC(シトシン)の90%の強度より大きいピークを、A(アデニン)のピーク候補として抽出する。

(5)

特開2002-168868

8

デニン) や T (チミン) より大きいピークを、C (シトシン) のピーク候補として抽出する。

【0029】 [2] ピーク間隔の検定

抽出したピークの前後の間隔を調べ、間隔が狭い2つのピーク、及び前後の間隔が共に広いピークは、候補から削除する。ただし、T (チミン) の移動度が遅いことを考慮し、T (チミン) の前後については、ピーク間隔の1/2程度のずれは許容する。さらに、同じ塩基が3連続以上する場合には、両端を除くピーク信号は優先的に残す。

【0030】 [3] 信号強度比率を計算し、マトリックス値を得る。

検定で残ったピークの信号強度比率を計算し、塩基毎に中央値を求め、その値を代表値とする。平均値ではなく、中央値にしているのは、ノイズの多い系では平均値は時々真値から外れた値を出すからである。そして、4種類の代表値から行列をつくり、その逆行列を求めてマトリックス値とする。

【0031】 [4] 信号波形をマトリックス変換して、ベースコードを行なう。

【5】ベースコードの結果から、さらに最適なマトリックス値を得る。ベースコードの結果、ほぼ確実に正解であるという重み付けをされた塩基(ピーク信号)を、全データ範囲について抽出し、その信号強度比率をもって、新たなマトリックス値を計算する。

【6】マトリックス値をファイルに保存

【0032】この泳動を行なったDNAシーケンシングユニットの認識番号と、ETターミネータの印を付加して、このマトリックス値をファイルに保存する。今後、*

*このユニットでETターミネータを泳動した場合には、ベースコードは自動的にこのマトリックス値を参照することになる。実施例のように、その系に対応した方法論のチューニングは、反応試薬キットや検出部を含む泳動系に大いに依存している。ある時には手順が前後したり、また全く逆の条件設定が必要となったりすることがあるかも知れない。しかし、状況に応じた適切な処置が必要であり、それがまた高速かつ精度の高いベースコードの条件ともなっている。

10 【0033】

【発明の効果】本発明では、蛍光色素ごとの検出部から得られた波形信号にマトリックス変換を行なって塩基ごとの信号波形を求め、それに基づいて塩基配列決定を行なう方法において、実際のサンプル泳動から得られるピークから所定の条件を満たすものを抽出し、それらのピークを用いてマトリックス値を得るようにしたので、専用試薬キットを用いなくても塩基配列を決定することができるようになる。

【図面の簡単な説明】

20 【図1】本発明の1つの局面を示すフローチャート図である。

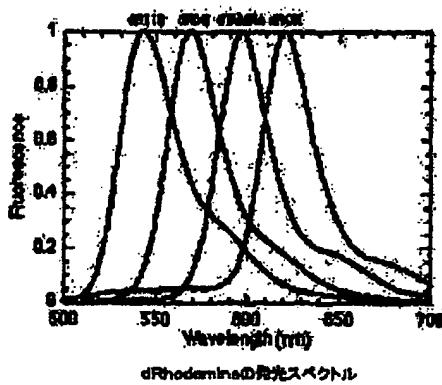
【図2】dRhodamin色素の規格化された発光スペクトルを示す図である。

【図3】各検出部で検出される塩基Aの信号ピーク波形を示す図である。

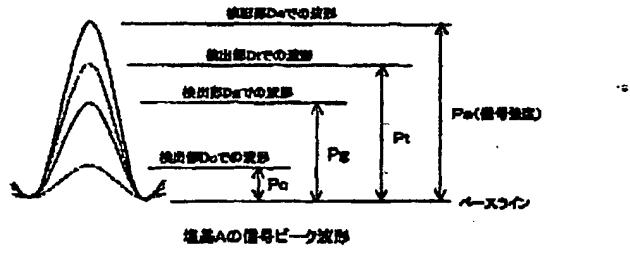
【図4】塩基により移動度が異なることによるピーク位置のシフトの様子を概略的に示す図である。

【図5】信号強度が逆転している場合を例示する図である。

【図2】



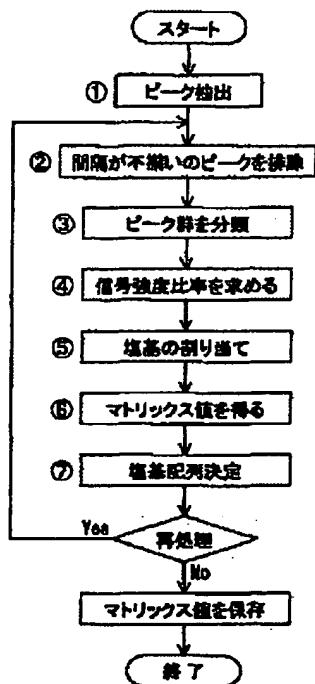
【図3】



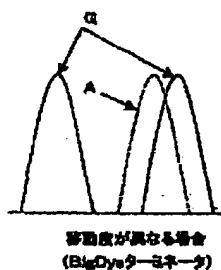
(6)

特開2002-168868

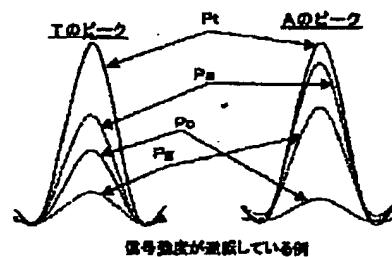
【図1】



【図4】



【図5】



フロントページの続き

F ターム(参考) 2G043 AA06 BA16 CA03 DA02 DA05
 EA01 EA19 FA06 JA01 KA02
 MA04 NA01 NA02
 2G045 AA36 DA12 DA13 FB07 FB12
 GC15
 4B024 AA20 CA01 HA19
 4B029 AA07 BB20 FA15
 4B063 QA13 QQ42 QR42 QR66 QS16
 QS34 QS36 QS39 QX02